Niveau(x) d'enseignement: 6 Numéro EDUQ : 0670372
Centre d'intérêt: 570

Gagné, François, 1940-.

A few remarks about the Evaluation Function within the PERFORMA Project / Francoys Gagné and Danielle Guillet ; transl. by Betty Howell. -- [Sainte-Foy] : INRS-Education, 1975.
8, [4] f. ; 28 cm.

Titre de la couv. Document: D-028. Bibliogr.: f. [9].

Vedette(s)-secondaire(s): Guillet, Danielle, 1953- * INRS-Education

Descripteur(s): Perfectionnement des enseignants * Programme de perfectionnement * Evaluation de programmes * Méthode d'évaluation * Universitaire (Ordre d'enseignement)

Identificateur(s) de nom: PERFORMA (Projet)

L'INRS-Education a collaboré en 1974-1975 au développement du projet PERFORMA, en prenant notamment la responsabilité de l'évaluation. En suivant un sentier qui n'a pas toujours été le plus direct, l'équipe chargée de l'évaluation a délimité ses tâches à l'intérieur du projet PERFORMA. Elle a dû aussi consacrer un temps considérable à clarifier son propre rôle et sa propre place à l'intérieur du projet. Le texte présente deux aspects distincts de ce processus de clarification. La première partie du texte montre comment on est arrivé à établir le cadre conceptuel le plus adapté au projet, et à définir la place et le rôle de l'équipe d'évaluation. La seconde partie du texte fournit une information plus précise sur les contenus et les méthodes d'évaluation.

Type de document: r

EDUQ 8670372

I N R S - Education

A few remarks about the Evaluation Function within the PERFORMA Project

Françoys Gagné
and
Danielle Guillet

Translated by Betty Howell

October 1975

Document D-028

As a result of an agreement made between the University of Sherbrooke and l'Institut National de la recherche scientifique (INRS), the Centre INRS-Education has collaborated since autumn 1974 on the development of the PERFORMA Project, taking particular responsibility for the EVALUATION tasks.

Following a path which has not always been the most direct, the evaluation team has progressively delineated its tasks within the PERFORMA Project. In addition to specifying the contents and methods of its initial concrete activities, the evaluation team has had to devote a considerable amount of time to clarifying its own role and its position within the project. That is why this descriptive text presents two distinct aspects of the clarification process. The first part of this process consisted of establishing the conceptual framework most suited to the project, as well as the position and the roles of the evaluation team; the second part was intended to give more precise information about the contents of the evaluation and the methods most appropriate to it.

By reason of the close interaction between these two aspects, the work to be done by the evaluation team will be specified with the least amount of conflict if its roles are clarified in advance.

In a context which is as changeable as that of a large-scale project in the process of evolution, and, what is more, with a team which has only been working full-time for three months, it goes without saying that the positions described in this text cannot by any means be considered final.

I - Some remarks on the clarification of roles

In order to illustrate the problems which this sector has posed, we have chosen two from among those we consider most crucial: the clear identification of the client and the precise nature of our role in view of the "formative - summative" role alternatives described by Scriven (1967). A short definition of our conceptual framework proceeds the two illustrations.

A. Conceptual framework

Worthen and Sanders (1973) divide the principal theoriticians of evaluation in education into three major currents:

a) "judgmental strategies" (Cronbach, Scriven, Stake);

b) "decision-management strategies" (Alkin, Stufflebeam);

c) "decision-objective strategies" (Hammond, Provus).

For various reasons, related both to the evaluators and to the context, our evaluation team oriented itself toward the school of thought represented by Stufflebeam and Alkin. The latter defines evaluation as follows:

> "...the process of ascertaining the decision areas of concern, selecting appropriate information and collecting and analysing information in order to report summary data useful to decision-makers in selecting among alternatives." (1969, p. 2)

In this context, evaluation becomes a "service" function for a decision-making group (the client), which will itself take on the definitive role of judge rather than delegating it to the evaluator.

B. The client of the evaluation team

The evaluation team has had to decide what its relationship would be to five decision-making groups, all of which could assume certain functions usually given to the client, especially the role of defining, with the evaluation team, those areas of priority decisions which guide the data collection. These five groups, in order of importance, are:

within the project:
- a) local respondants in the colleges;
- b) PERFORMA administration;

outside the project:
- c) administration of the University of Sherbrooke;
- d) the groups which subsidize educational research (e.g. Canada Council, FCAC);
- e) the Quebec Ministry of Education, here represented by the Direction générale de l'enseignement supérieur (DGES).

We have discussed elsewhere at greater length (Gagné, 1975) the problems associated with the amgiguity we have experienced (and which has not been completely resolved) in identifying the primary client in terms which would be clear and acceptable to all possible "clients" with which the evaluation team deals. The fact that it is clear to us and to the PERFORMA administration that they are the primary client still does not prevent other direct or indirect, voluntary or involuntary, pressures from being exerted in the future by the other decision-making groups mentioned above to influence the roles and activities of the evaluation team.

C. The "summative" and "formative" concepts

These two terms, invented by Scriven (1967), have been adopted by many other authors, particularly Stake (1967) and Bloom (1971). In order to avoid the confusion which frequently surrounds their use, let us recall that in their original sense they refer to different roles played by the evaluators, rather than to differences in approach or in the contents of evaluation. These roles differ, not only in the timing of the evaluator's intervention (formative = project in the process of being developed or tried out; summative = completed product), but also in the nature of the client who is called upon to make decisions (formative = project designers; summative = outside administrators or users).

It follows therefore that clarification of the identity of the primary client by the evaluation team (cf. section B above) as well as the experimental nature of the project necessarily lead to the adoption of a formative role. Several concrete corollaries of this role choice are:

a) analyses comparing this project to other, similar ones become a secondary concern;

b) collection of data for use in improving the project becomes the primary concern;

c) there is at least as much emphasis on the methods of operation as on the anticipated effects;

d) short-term rather than long-term operations are favored.

II - Some remarks on the clarification of the contents

Recently, the evaluation team made a rather extensive inventory of problems or questions raised during the course of the preceeding months from documentation and encounters with various members of the PERFORMA project. To give some structure to this set of data, we have set up a matrix, a simplified version of which is presented in Table 1.

The horizontal axis (columns), essentially inspired by the evaluation model proposed by Alkin (1969), represents the five areas in the development of an educational program where evaluation activities may intervene to assist in the decision-making process. We have attached a copy of the short description which the author himself has made of these five evaluation need areas (cf. appendix).

The vertical axis (rows) corresponds to a summary division of the different parts of the C.P.E.C. (Certificat de perfectionnement en enseignement collégial)[1]. The C.P.E.C. is a set of means (Curriculum,

---

[1] C.P.E.C. can be translated approximately as "Certificate of Further Development in College Teaching".

Table 1

Simplified classificatory matrix of evaluation contents

	PLANIFICATION		EXPERIMENTATION		CERTIFICATION
	Systems Assessment 01	Program Planning 02	Implementation 03	Improvement 04	05
(Who) Clientele 10	11	12	13	14	15
(What) Curriculum 20	21	22	23	24	25
(How) 30 A. Instructional Design	31	32	33	34	35
B. Administrative Structures 40	41	42	43	44	45
(Where) Environment 50	51	52	53	54	55
(Why) Results 60	61	62	63	64	65

Instructional and Administrative Structures), arranged in an Environment and addressed to a specified Clientele, whose objectives are operationalized in the form of expected Results.

We have been able to place the main questions or problems encountered in one or another of these cells. We believed that this schematization would be of assistance in the task of identifying priorities, which must be done in close collaboration with the client. This "negotiation" is now under way.

A. General problems

Before illustrating the contents of certain of these cells, let us take a brief look at the use we made of the horizontal and vertical coordinates to describe our priority evaluation tasks, starting from the spontaneous observation that even a rather superficial examination of all the questions raised would go well beyond the limits of available resources. The following observations are related to one or another of the colums or rows taken as a whole rather than to one of the individual cells.

1. Horizontal axis (columns 01, 02...)

a) The contents of columns 03 and 04 (Implementation and Improvement) are the most familiar to us. Here we find the traditional

vement) ought to take precedence over the others, since the C.P.E.C. is now in an experimental phase.

b) At first glance, columns 01 and 02 (Systems Assessment and Program Planning), which refer to already completed stages, might be considered irrelevant. However, they should not be completely ignored. For example, at the Implementation stage, the evaluation team will doubtless face certain deficiencies, inaccuracies or omissions carried over from the Program Planning stage. In short, certain "a posteriori" evaluative activities may turn out to be inevitable.

c) Column 05 (Certification) could be considered the most important, since the C.P.E.C. will complete its three-year experimental phase next June. However, from the standpoint of the evaluation team, it is irrelevant. First of all, it calls directly for a "summative" role, one which we have flatly rejected (cf. section A.3 above). Moreover, the implementation delays shortened the effective experimental period in such a way that the overall judgment envisaged for the Certification stage seems to us at this moment to be premature. Unless major problems should arise, which would make the very existence of the program questionable, it would appear necessary instead to envisage a possible extension of the experimental phase.

2. Vertical axis (rows 10, 20...)

Except for the rather marginal nature (in our view) of the row entitled "Environment", at least in so far as it concerns evaluation activities, the vertical axis poses only one major problem, related to the results of the program.

The goal of teacher training programs like the C.P.E.C. is to provoke changes in attitudes and behavior in the teachers in a direction considered desirable. The desirability of changes is usually based on the "teaching effectiveness" of the new, expected behavior. In other words, the ultimate criterion for judging the greater or lesser success of the C.P.E.C. program would consist in showing that the students whose teachers have completed a C.P.E.C. program succeed better because of changes in attitudes and/or behavior on the part of these teachers, changes clearly attributable to the training program.

A partial review of the scientific literature on this subject does not encourage us to believe that such a demonstration is possible. Thus, a very recent survey by Turner (1975) of research on the impact of teacher training on subsequent success in teaching, comes to the following conclusion:

> "In spite of recent improvements in research in the field, the amount of dependable information available compared to the amount needed to formulate more effective policies and practices in teacher education is miniscule (sic)." (p.107)

In a related area, that of research done to associate certain kinds of teaching behavior with gains in learning by students, Rosenshine and Furst (1971) begin their text with this dramatic affirmation:

> "This review is an admission that we know very little about the relationship between classroom behavior and student gains. It is a plea for more research on teaching. It is also a plea to educational researchers and to teacher educators to devote more time and money to the study of classroom teaching." (p. 37)

Finally, in another closely related sector, the study of the relationship between two distinct criteria of teaching success, i.e. the judgments of the students and gains in learning achieved, Doyle (1975) systematically reviews the results of research in this area. He concludes as follows:

> "There is no acceptable basis in these studies for accepting the hypothesis of a negative relationship between ratings of over-all instructor ability or effectiveness and learning-oriented criterion measures. On the other hand, neither are the data persuasive that there is a very large positive relationship. There does, however, appear to be a fairly consistent, low-to-moderate positive correlation between general ratings and student learning." (p. 65)

We might add to this group a Quebec study (Baby and others, 1973) which proposed to evaluate the differences in attitudes and behavior between graduates of an experimental teachers' training program and graduates of "traditional" institutions. While several differences (rather minor in our view) did appear at the level of attitudes, measurements made of behavior showed no differentiated profile.

In short, this accumulation of negative or only slightly significant results, often after the investment of considerable effort, has led us to put aside a planned project which was specifically intended to measure changes in teaching attitude and behavior induced by the program. This decision was also based on serious methodological problems which added to the arguments taken from the literature.

B. Outline of activities

Even though the "negotiation" of priorities has not yet been concluded, the evaluation team has been able to clarify enough matters with its client about the relative importance of the various cells of the matrix to be able to undertake certain data-gathering operations and start on the structurization of other analyses.

1. Projects under way

Two studies are already being carried out: one on the socio-professional characteristics and modes of progress of the clientele, the other on a detailed analysis of the instructional design of the activities in the training program.

a) Study of the clientele

Because the C.P.E.C. program is offered on a voluntary basis, it is quite possible that the clientele effectively reached is not representative of the possible population as regards various characteristics (age, experience, subject taught, etc.). The identification of a sub-population with special characteristics might suggest certain preliminary hypotheses for an eventual survey of those not in the program (cf. cells 13-14 in Table 1).

The second aspect of this study concerns the different aspects of the teachers' progress through the C.P.E.C. program. For example, what is the average rate of progess in terms of accumulation of credits (cells 43-44)? Are "hibernation" phenomena - i.e. prolonged interruptions in participation - observable (cell 44)? Do some teachers tend to over-specialize in a certain subject when choosing activities (cell 23)? To what extent are the routes taken individualized (cell 23)? Etc...

We have extracted most of the information needed to do both phases of this study from the teachers' application files.

b) Analysis of the curriculum and the instructional structure

Scriven (1967) distinguished between two types of evaluation, which he called respectively "pay-off" and "intrinsic" evaluation. In the former, the evaluation is made of the results produced, while in the latter the emphasis is on the analysis of intrinsic qualities of the program. In our case, since we have put aside, at least temporarily, any attempt to obtain an objective measurement of the main effect of the program (cf. pages 5-6), which is the acquisition of more effective teaching attitudes and behavior by the teachers taking part in the C.P.E.C. program, we are led to concentrate almost exclusively on an intrinsic evaluation. An extensive analysis of the C.P.E.C.'s curriculum and instructional structure therefore assumes a prime importance, since these two aspects are the essence of what is done to achieve the intended goals.

In order to accomplish this task, we have constructed a detailed descriptive chart of the activities, which must be completed by those responsible for an activity (Gagné and Guillet, 1975). The seventy (70) or so questions it contains cover the main stages of these teaching interventions, notably: the types of needs to which the activity responds (cells 13-14), the importance of the participation of

teacher-students in setting up the activity (cells 33-34), the clientele reached (cells 13-14), the different teaching methods used (cells 33-34) the characteristics of the human resources (cells 43-44), the roles played by these resources (cells 33-34), the use of personal experiences and/or teaching materials from the teachers' classrooms (cells 33-34), the proximity of administrative control exerted by the PERFORMA administration (cells 43-44), the modes of evaluating learning (cells 33-34), etc...

Moreover, various rearrangements of the collected data might permit the presence of institutional profiles to be verified, as well as a longitudinal evolution in certain components of the instructional design.

2. Future projects

Some other projects are presently undergoing preliminary study; they will soon be examined in more detail by the evaluation team and the PERFORMA administration. Strictly for illustrative purposes, however, a brief description of some of the themes taken from our inventory will be found below.

a) The development of a questionnaire to permit the teacher-students to express in a more structured manner than at present their opinions on the main aspects of the activity they have just completed (cells 24-34).

b) A study of a sample of teacher-students intended to clarify among other things the needs to which the C.P.E.C. responds, transfer noted between the training activities and their daily teaching duties, the (subjective!) changes noted in their teaching attitudes and behavior that they would attribute to the C.P.E.C. activities (cells 14, 34, 64).

c) A study of non-participating teachers to clarify the reasons they do not participate, in order to make a more precise estimate of the probable clientele of the C.P.E.C. program in an institution (cell 14).

d) A more thorough evaluation, on the basis of preliminary data (cf. b) above) of whether an objective measure of changes of attitude and behavior, said to be attributable to the C.P.E.C. activities, can be obtained.

o - o - o

References

ALKIN, M.C. (1969) Evaluation Theory Development. Evaluation Comment, v. 2, no. 1, pp. 2-7.

BABY, A. et autres (1973) Projet CEXFORME: rapport d'évaluation. Tomes I et II. Québec: Faculté des Sciences de l'Education, Université Laval.

BLOOM, B.S., HASTINGS, J.T. and MADAUS, G.F. (1971) Handbook on Formative and Summative Evaluation of Student Learning. New York: McGraw-Hill.

DOYLE, K.O. (1975) Student Evaluation of Instruction. Lexington: D.C. Heath.

GAGNE, F. (1975) Le client de l'équipe d'évaluation de Performa. INRS-Education: document non publié.

** ROSENSHINE, B. et FURST, N. (1971) Research on Teacher Performance Criteria. In Smith, B.O. (Ed.) Research in Teacher Education: a Symposium. Englewood Cliffs: Prentice-Hall, pp.37-72.

SCRIVEN, M. (1967) The methodology of evaluation. In R.E. Stake (Ed.) Curriculum Evaluation. AERA monograph series on evaluation, no. 1. Chicago: Rand McNally.

STAKE, R.E. (1967) The countenance of educational evaluation. Teachers College Record, v. 68, pp. 523-540.

TURNER, R.L. (1975) An Overview of Research in Teacher Education. In Ryan, K. (Ed.) Teacher Education. NSSE: 74th Yearbook. Chicago: U. of Chicago Press, pp. 87-110.

WORTHEN, B.R. et SANDERS, J.R. (1973) Educational Evaluation: Theory and Practice. Worthington: C.A. Jones Publishing.

** GAGNE, F. et GUILLET, D. (1975) Grille d'analyse des activités du CIEC. Québec: INRS-Education, document D - 14.

Appendix A

The contents of this appendix are taken

from pages 3 to 5 in M.C. Alkin's (1969) article

EVALUATION NEED AREAS

The foregoing definition and assumptions are closely tied to the decision-making process, which in turn leads to a consideration of what kinds of educational decisions require evaluative information. Inquiry along these lines has led to the development of a decision-oriented classification of the various types of evaluation. Five areas of evaluation may be identified.

These five areas represent attempts to provide evaluative information to satisfy unique decision categories. In other words, there are evaluations necessary in providing information for decisions about the state of the system. (We call such evaluations *systems assessment.*) There are evaluations necessary in providing information to assist in the selection of particular programs likely to be effective in meeting specific educational needs. (We call this kind of evaluation, which takes place prior to the implementation of the program, *program planning.*) There are evaluations necessary in providing information relative to the extent to which a program has been introduced in the manner in which it was intended and to the group for which it was intended (*program implementation*). There are evaluations necessary in providing information during the course of a program about the manner in which the program is functioning, enroute objectives are being achieved, and what unanticipated outcomes are being produced. Such information can be of value in modifying the program (*program improvement*). Evaluations are necessary in providing information that might be used by decision-makers in making judgments about the worth of the program and its potential generalizability to other related situations (*program certification*).

The evaluation areas outlined above seem to represent a growing consensus among a number of people engaged in the study of evaluation. The first two and the last need areas discussed are somewhat similar, respectively, to "context," "input," and "product" presented by Stufflebeam (1966). What he refers to as "process" we have chosen to think of as two separate stages, program implementation and improvement. Major differences of emphasis are found between the descriptions of our need areas and his stages. However, Stufflebeam's work contributed substantially to our thinking. The discrepancy model presented by Provus (1969) outlines five stages: "definition," "installation," "process," "product," and "cost benefit analysis." The first four of these are somewhat similar to our stages two through five. Provus does not include systems assessment as a part of his model. With respect to cost benefit analysis, we would maintain that cost benefit considerations are a part of each need area of the evaluation and cannot be thought of as simply an additional task to be attached when all of the other evaluations have been completed. The general notion that cost benefit considerations are a part of each stage of the evaluation process also would seem to be subscribed to by [illegible] Hemphill in his recent NSSE Yearbook chapter (1969). The evaluation types outlined by Rodney Skager (1969) are quite similar to the framework presented here. A better understanding of the need areas will be gained by a more complete description of each.

Systems Assessment.

Systems assessment is a means of determining the [illegible] and specificity of educational objectives appropriate for a particular situation. The needs may be represented as a gap between the goal and the present state of affairs. The evaluative problem, then, becomes one of assessing the needs of students, of the community, and of society in relation to the existing situation. Assessment, therefore, is a statement of the status of the system as it presently exists in comparison to desired outputs or stated needs of the system.

A systems assessment might be related to evaluation of a specific instructional program and thus the charge would be to determine the present status relative only to a specific objective and related objectives. We would refer to this as a "sub-system assessment."

Systems assessment does not refer to specification of process characteristics appropriate for a district, school, or classroom. A statement such as "this district needs a low pupil-teacher ratio" or "a need of this district is to institute team teaching" is not a systems assessment. The system assessment must be related to the ultimate behavior of clients of one type or another (pupils, parents, community, etc. – all clients of the school). To put it simply, system assessment must result in a statement of objectives in terms of outputs of the school.

The process in the systems assessment area of ascertaining the decision area, specifying and collecting information and reporting summary data, requires methodology and techniques different from that which might be employed, for example, in a typical experimental design. The data are concerned with the status of the system. The summary data might be comparative, historical, or other descriptive information.

Program Planning

Program planning, the second need area, is concerned with providing information which will enable the decision-maker to make planning decisions – to select among alternative processes in order to make a judgment as to which of them should be introduced into the system to fill needs efficiently the critical needs previously determined. In an instance where we are proceeding through severe needs areas in sequential fashion, the following might occur. After the decision-maker receives the systems assessment evaluation, he might make a decision as to the appropriate means of fulfilling that need. Alternatively, he might designate several possibilities and ask the evaluator to provide information on the possible impact of each. Hence, in program planning, the evaluator provides the data for an evaluation of a program prior to its inception. The task of the evaluator is to anticipate the attainment of goals and to assess the potential relative effectiveness of different courses of action.

It is quite obvious that the collection and analysis of data of the type required for this evaluation need area will be quite different from collection and analysis problems for other areas. The techniques may require both internal and external evaluation procedures. (See Lumsdaine, 1965.)

By way of internal evaluation, programs may be examined to determine the extent to which their reproducible segments purport to achieve the objectives of the program being evaluated. Technical features of style or construction, practicality and cost are other means of providing internal evaluation. To date, the evaluations of products by EPIE (see p. 11), have been primarily based upon internal evaluations.

External evaluations of programs yet to be implemented might take the form of examining research data on the results of implementation in similar or near-similar situations. Or external evaluations might attempt to utilize some of the various educational planning techniques to obtain data. Computer simulations might be developed; Delphi analysis might provide insights into the potential outcomes of a program; gaming and various other systems analytic approaches might also provide external evaluation data.

Program Implementation

After the decision-maker has selected the program to be implemented, an evaluation of program implementation determines the extent to which the implemented program meets the description formulated in the program planning decision. In the case of an existing program where no known changes have been implemented, the evaluation task at this stage is to determine the degree to which planning descriptions of the program coincide with the implemented program and the extent to which assumed descriptions of inputs to the system (students) correspond with observed inputs.

There have been numerous examples in the educational literature of conflicting results relative to the impact of a specific instructional treatment. We would maintain that in large part this is attributable to the lack of specificity of the precise nature of the instructional treatment that was employed. Team teaching is *not* always team teaching. More precisely, team teaching in Santa Rosa might be quite different from team teaching in Boston or from team teaching in Palo Alto, California. The precise definition of the parameters defined as team teaching in a given situation would help to insure an understanding of what is being evaluated and whether what is being evaluated is what the investigator thought the program was.

Program Improvement

The evaluator can play an important role in program improvement, the fourth need area, by providing as much information as possible about the relative success of the parts of the program. In order to perform program improvement evaluation, it is necessary to recognize the basically interventionist role that the evaluator has been asked to take.

The key point in the understanding of the role of the evaluator in performing evaluations in this need area is that he is first and foremost an interventionist attempting to provide data which will lead to the immediate modification and, hopefully, improvement of the program. As the evaluator identifies problems and collects and analyzes related information, data are presented immediately to the decision-maker so that changes may be executed within the system to improve the operation of the program. Information might include data on the extent to which the program appears to be achieving the prescribed objectives, as measured by regular tests; information also might be presented which relates to the impact of the program on other processes or programs.

This need area has often been overlooked or ignored by the traditional evaluator who has attempted to impose the antiseptic sterility of the laboratory on the real world. Such an approach may make for a fine experiment, but it does little to improve a program which is often not in its final form.

Program Certification

In the fifth evaluation need area, program certification, the role of the evaluator is to provide the decision-maker with information that will enable him to make decisions about the program as a whole and its potential generalizability to other situations. The evaluator might attempt to provide information which will enable the decision-maker to determine whether the program should be eliminated, modified, retained, or introduced more widely.

The kind of information collected for program certification decisions is in large part dependent upon who is the intended decision-maker. It is obvious that different information will be required if the potential decision-maker is the teacher, the principal, or a funding agency. Evaluations in this area will be concerned with examining the extent to which the objectives have been achieved, as well as with the impact on the outcomes of other programs.

In program certification evaluations, there is a requirement for valid and reliable data which would generally require that the evaluator attempt to apply as rigid a set of controls as possible. The evaluator might use pre- and post-test designs and employ sophisticated methods for analyzing the data. Intervention should be avoided in evaluations in this need area. Here the traditional evaluator is "at home."